

Evaluation Report on the Nebraska State Department of Education's District Assessment Portfolio Training Process

Ellen Forte Fast, Ph.D.
Draft—October 11, 2004

Overview

This evaluation report on the Nebraska Department of Education's District Assessment Portfolio training process addresses the following questions:

- ♦ How clearly did the Nebraska Department of Education communicate its expectations for the District Assessment Portfolios to the local educators who prepared these portfolios?
- ♦ How did the criteria for scoring the District Assessment Portfolios align with the criteria that were communicated to local educators?

In addition to addressing these questions, this report includes recommendations for future refinements in the District Assessment Portfolio training process.

Background

Nebraska is unique among the US states in its approach to assessing students' academic knowledge and skills. Under the *No Child Left Behind Act of 2001* (NCLB), all states are required to assess students' knowledge and skills in relation to academic content standards in reading/language arts and mathematics in grades three through eight and at the high school level. But, while all other states use the same assessment statewide in at least some grades and content areas¹, Nebraska associates the responsibility for assessment with the responsibility for instruction and bases control for both at the local level. This perspective is evident even in the name of its system for supporting local assessment and reporting: the School-based Teacher-led Assessment and Reporting System (STARS). As Nebraska's Commissioner of Education, Dr. Doug Christensen, notes in the forward to a publication describing STARS, this system "is firmly grounded in the belief that decisions about student learning should be standards-based and should be based upon classroom knowledge of the student" (STARS Summary, June 2002. p. i).

Placing the responsibility for assessment at the local level does not, however, represent a ceding by the Nebraska Department of Education (NDE) of control for the quality of the assessments. Rather, the role that the NDE plays in the assessment system is one of guidance and support. This role is noted in the opening statement of the STARS document noted above:

The underlying philosophy that supports Nebraska's School-based Teacher-led Assessment and Reporting System emphasizes a partnership between the local school districts and the Nebraska Department of Education. Keeping decisions about student performance on standards at the local classroom level provides a balance between state level guidance and local decision-making. Partnership and balance are the two crucial elements in making changes in schools that will result in improved learning for all students. (STARS Summary, June 2002, p. 1)

¹ Although Iowa technically does not operate a statewide assessment system, nearly all local school districts have adopted the Iowa Tests of Basic Skills and the Iowa Tests of Educational Development. Thus, these assessments represent de facto statewide tests.

The state's responsibility for ensuring the quality of local assessments is also built into the federal education policy as most recently reauthorized in NCLB. Under Section 1111(b)(5) of this law, states that allow for the local development and/or adoption of standards and assessments are required to ensure that these standards and assessments meet the same criteria for quality expected of statewide assessments.

The NDE's own perspective and the federal requirements both place the ultimate responsibility for assessment quality with the state. This means that the NDE must implement a system for ensuring assessment quality. The system that the NDE has developed for this purpose involves:

- ♦ establishing criteria for assessment quality;
- ♦ structuring the format in which local educators are to provide evidence in relation to these criteria (i.e., the District Assessment Portfolios);
- ♦ communicating these criteria and evidence structure to local educators who are responsible for creating the District Assessment Portfolios; and
- ♦ scoring the District Assessment Portfolios in a manner that is consistent with the criteria for assessment quality as communicated to local educators.

As part of this system, the NDE has contracted with independent outside evaluators to review and analyze the system's components and processes. The evaluation summarized in this report focused on the latter two bullets listed above: the communication of the assessment quality criteria and the process for scoring the District Assessment Portfolios.

It is important to note that this evaluation addresses the *processes* for communication and *processes* for scoring. This evaluation does not represent a review of the STARS system in general, a critique of the assessment quality criteria themselves, or a review of the *results* of the portfolio scoring process. Recommendations regarding some of these issues are included in the final section of this report, but these are incidental to the main focus of the evaluation.

Further, this evaluation is limited to the communication and scoring processes that occurred at the end of the 2003-04 school year. Documents from prior school years were reviewed only if these were relevant to the 2003-04 District Assessment Portfolio submissions. The evaluator did not attend or review evidence of training or scoring sessions for prior years.

Evaluation Process and Results

As indicated in the Overview, this evaluation was intended to address the following questions:

- ♦ How clearly did the Nebraska Department of Education communicate its expectations for the District Assessment Portfolios to the local educators who prepared these portfolios?
- ♦ How did the criteria for scoring the District Assessment Portfolios align with the criteria that were communicated to local educators?

The NDE hired an independent external evaluator to gather evidence related to these questions and to provide this evaluation report summarizing an analysis of this evidence. This evaluator has extensive experience with state assessment systems and has served as a peer reviewer of state assessment and accountability systems on behalf of the US Department of Education for several years. She had no prior direct experience with the NDE or its assessment systems.

The remainder of this section summarizes the evidence related to the evaluation questions.

Question 1: Communication of the Quality Assessment Criteria

The first question addressed in the evaluation involves the clarity with which the NDE communicated its expectations for assessment quality to the local educators charged with creating the District Assessment Portfolios.

Two types of evidence were considered in relation to this question. First, the evaluator reviewed copies of the documentation regarding the District Assessment Portfolios that the NDE disseminated to local school districts. Second, the evaluator attended portfolio pre-submission workshops for local educators designed to support the development of the local District Assessment Portfolios and gathered evidence of the content of these sessions.

Documentation Review. In May of 2004, the NDE provided the following documents for review as part of this evaluation:

1. STARS: A Summary (June 2002)
2. STARS Update # 14 (March 2004)
3. Assessment Portfolio Instructions and Suggestions—Mathematics 2003-04 (March 2004)
4. A blank District Assessment Portfolio submission form
5. Blank sample charts for the submission of specific evidence in the District Assessment Portfolios
6. Sample District Assessment Portfolio for reading, listening, and speaking and for mathematics

All of these documents had been provided to local educators prior to or during the 2003-04 school year. For each document, the purpose, contents, and comments are indicated in Table 1.

Each of these documents provides information related to the preparation and submission of the District Assessment Portfolios. In total, they represent a comprehensive set of documents and forms related to the purpose of the District Assessment Portfolios, what evidence they must include, how they may be structured, and how they will be scored. This information is grounded in the larger context of how the assessment results are to be used as part of a standards-based reform system.

The six quality criteria for local assessments are described repeatedly in this set of documents. The rubric used to score the portfolios, which lists the characteristics associated with the three portfolio ratings for each criterion (Not Met, Needs Improvement, and Met), is also presented several times. In each case, the same language is used to define criteria and the same examples are used to exemplify the evidence districts must submit. The narratives in each of these documents are relatively brief and to-the-point.

As illustrated in the timeline below, most of these documents were released in March of 2004, approximately three months prior to the pre-portfolio workshops and nearly four months prior to the portfolio submission deadline.

Timeline for the release of supporting documents							Update #14
Documents are listed above the month of release							Instructions
STARS: A Summary							Sample charts
▼							Sample portfolios
June	Sept	Dec	March	June	Sept	Dec	▼ March June
	2002-03 school year				2003-04 school year		Portfolios due ►

Table 1. Documents Provided to Local Educators Prior to the Submission of the 2003-04 District Assessment Portfolios

Document and Purpose	Contents	Comments
STARS: A Summary (June 2002) Purpose: To describe the purpose and components of STARS	I. STARS Purpose and Philosophy II. Description of the standards, assessment, and accountability components of STARS III. State and local responsibilities for STARS IV. How STARS data are analyzed, reported, and used to support school improvement Appendices: - District Assessment Portfolio Rubric - Quality Criteria Rating Chart - Scoring guides for Writing Assessment	This document provides a concise summary of why STARS was developed, what it encompasses, and how the information is used. The District Assessment Portfolios are explained in the context of the full system, which may support local educators' understanding of why the NDE required districts to submit portfolios.
STARS Update # 14 (March 2004) Purpose: To provide local educators with current information	List of previously released documents Timeline for STARS-related activities during the 2003-04 school year I. Reporting Assessment Results in 2004 II. STARS & NCLB III. Preparing the School District Assessment Portfolio IV. Standards Update V. Assessment Update for Long-Range Planning (2006-07) Attachments: - District Assessment Portfolio criteria, rubric, and rating chart to be used through 2005-06 - Content standards for science, social studies, reading, and mathematics - District Assessment Portfolio rubric that will take effect in 2006-07	This document summarizes a broad range of information related to the STARS for the 2003-04 school year. Each topic is presented in a separate chapter, which provides organizational structure for the extensive information. Explaining the links between STARS and NCLB requirements, as well as the standards information, provides critical context for local educators. Information on changes to the District Assessment Portfolio rubric and other long-range activities can help districts to plan, but also represents transparency and support for change in the system. Changes are not introduced suddenly, without adequate time for district action. Rather, the NDE alerts districts to upcoming changes far in advance and also makes clear that it is monitoring and working to improve its own systems. The list of updates indicates that the NDE takes its responsibility for communicating its expectations seriously and regularly provides STARS information to local educators.
Assessment Portfolio Instructions and Suggestions—Mathematics 2003-04 (March 2004) Purpose: To delineate the District Assessment Portfolio submission requirements and to provide suggestions for the content and format of these submissions	I. General Instructions II. Suggestions for District Portfolio Development III. How Your Portfolio will be Evaluated	This document represents a step-by-step guide for developing the District Assessment Portfolio submission. Information on the portfolio review process is included so that local educators know how their portfolios will be evaluated. This encourages those developing the portfolios to consider the "audience", perhaps helping them to explain their evidence more clearly.

Table 1, continued.

Document and Purpose	Contents	Comments
Blank District Assessment Portfolio submission form Purpose: To provide local educators with the actual forms necessary for portfolio submission as well as a checklist to ensure that submissions are complete	District Assessment Portfolio Submission forms Self-Checklist	The NDE provides these forms to help prevent the submission of incomplete or otherwise inadequate portfolios.
Blank Sample Charts Purpose: To provide local educators with examples of how some types of evidence could be presented to ensure completeness and clarity	Criterion 1—matrix of assessments by standard, type, #of items, difficulty levels of items, and comments/changes Criterion 2—matrix of standards by assessment type, start date for instruction, end date for instruction, assessment date, comparison of instruction and assessment dates, and comments/changes; matrix of standards by month to record instruction Criterion 3— matrix of assessment by standard, test type, #of biased items, and actions taken Criterion 4— matrix of assessments by standard, test type, #of items, #of items considered developmentally appropriate, #of items requiring revision, and actions taken Criterion 5—matrix of assessments by standards, test type, #of items, #of students assessed, reliability, method used to estimate reliability, and comments; contrasting groups worksheet Criterion 6—matrix of standards by teacher judgment, actual results, and necessary changes; matrix of assessments by standards by test type, # of items, maximum possible score, item weighting, cut score setting process and outcomes	These sample charts provide examples of what specific information should be included as evidence related to each quality criterion.
Sample District Assessment Portfolios Purpose: To provide local educators with examples of successful submissions for use as guides in the development of their own submissions	Sample District Assessment Portfolios submissions for reading, listening, and speaking and for mathematics	These examples could help local educators to understand what types of evidence should be included in the portfolios and how this evidence could be presented to facilitate their review.

Portfolio Pre-Submission Workshops. The NDE conducted three regional portfolio pre-submission workshops prior to the June 30th submission deadline for the 2003-04 portfolios. As indicated on the agenda, these workshops were intended to:

- ♦ Reinforce and clarify the scoring criteria for District Assessment Portfolios;
- ♦ Connect the District Assessment Portfolio Scoring Rubric and the Self-Analysis Checklist;
- ♦ Illustrate varying performance levels of the District Assessment Portfolio criteria;
- ♦ Offer the opportunity for district teams to self-assess their District Assessment Portfolio documentation; and
- ♦ Offer the opportunity for district teams to discuss their District Assessment Portfolio documentation with peers.

These workshops were held on May 28th in Lincoln, on June 9th in North Platte, and June 11th in Kearney. The team of facilitators was essentially the same for all workshops and included state, regional, and district personnel. Sue Graupner of Lincoln Public Schools served as the lead facilitator; Tammy Heflebower (ESU 6), Mitzi Hoback (ESU 4), and Jan Hoegh (Aurora Public Schools) led specific sections of the workshops. The evaluator observed the Lincoln and Kearney workshops. The presentations and basic discussions were essentially the same across these two workshops.

District personnel responsible for developing the District Assessment Portfolios were encouraged, but not required, to attend a workshop. In most cases, participating districts were represented by several staff members. These district teams sat and worked together during the workshops.

Each workshop began with a brief opening period during which the facilitators introduced themselves and provided an overview of the agenda. The rest of the day was six periods, plus time for breaks and lunch, to coincide with the six quality criteria. During each of these six periods, a quality criterion was introduced and described briefly, the evidence requirements were described in some detail, and the participants were given time to ask questions and review their own evidence along with their teams.

Facilitators provided color-coded supporting documents for each criterion. These documents included a statement of the criterion, questions that should be addressed in the submissions, characteristics of each level in the portfolio scoring rubric, and exemplars of responses for each of these levels. Facilitators walked participants through each document. The exemplars were used in an interactive exercise in which the participants were asked to play the part of reviewers and determine why the samples warranted the scores they had been given.

In terms of basic structure and content, these documents were consistent with the detailed documents that had been released in March of 2004. Specifically, the STARS Update #14, the Assessment Portfolio Instructions and Suggestions, and the documents used in the workshops were organized such that each criterion was presented in turn, followed by a description of the portfolio scoring rubric. Within the section for each criterion, each of these documents was organized around the same three basic questions:

- (1) Who did the process?
- (2) What did they do?
- (3) What were the results?

Facilitators referred to these questions repeatedly during the workshops and used them to make connections between specific points in the scoring rubrics and specific statements in the exemplars. The

critical features of the necessary evidence for each of these questions are indicated in Table 2, along with issues that arose during the discussions of these questions.

Table 2. Evidence Required for Each Basic Question

Question	Critical evidence features
Who did the process?	Clearly described qualifications Independence of reviewers from developers
What did they do?	Clearly described process Appropriate and adequate process
What were the results?	Clearly described results Adequacy of results Adequacy of actions to correct any gaps or weaknesses

In general, participants seemed to understand the requirements for criteria 1, 2, and 4. Discussions related to criterion 3 (freedom from bias and insensitive or unfair situations) seemed to indicate some confusion between bias issues and basic editing issues. In the workshop documents, the exemplar for the Met performance level of this criterion includes a table with comments that are related to editorial issues, not bias issues. This may have contributed to some confusion on the part of participants regarding what the bias reviews were to address.

Workshop participants clearly found criteria 5 and 6 to be the most challenging to address. Identification of an appropriate method for estimating reliability (criterion 5) was somewhat problematic, and there was some confusion regarding the decision-consistency method described under criterion 5 and the contrasting groups method described under criterion 6.

Each workshop ended with a period for general questions and discussion. Some participants remained after the workshops to continue working with their portfolio development teams.

Evaluation Results. The evidence described above was considered in relation to the following question:

How clearly did the Nebraska Department of Education communicate its expectations for the District Assessment Portfolios to the local educators who prepared these portfolios?

Through the documents and the workshops, it appears that the NDE communicated its expectations for the District Assessment Portfolios relatively clearly and consistently. The same information was presented multiple times through documents and during the pre-submission workshops. A local educator who read these documents and attended a workshop would not have gotten contradictory information. Rather, the workshops reinforced and further clarified the information that had already been published by the NDE.

Given the confusion apparent in the workshops, the descriptions of expectations for criteria 3, 5, and 6 may need to be clarified.

Recommendations. The following recommendations may contribute to the continued improvement of the portfolio submission process.

- (1) Continue to offer the portfolio pre-submission workshops.

- (2) To the extent possible, engage the same facilitators who led the 2004 workshops for workshops in subsequent years.
- (3) Involve some experienced scorers in the portfolio pre-submission workshops to provide guidance and to support local educators' understanding of their audience when developing the portfolios.
- (4) Provide the Instructions and Suggestions document, sample charts, and color-coded documents to districts at the beginning of the school year.
- (5) For the sample portfolios, consider using the same responses that are used in the exemplars presented in the workshops and annotate these samples (e.g., use callouts) to point out the critical features for each criterion and question.
- (6) Continue to refine the Instructions and Suggestions document based on the feedback and questions from participants in the workshops.
- (7) Clarify differences between bias issues and editorial issues in the documentation for criterion 3.
- (8) Continue to seek practical and technically-sound solutions for estimating reliability and setting performance standards for the locally-developed assessments.

Question 2: Consistency between Pre-Portfolio Training and Training for Portfolio Scoring

The second question addressed in this evaluation involves the consistency between the portfolio information that the NDE communicated to local educators and the process and criteria used to score the portfolios.

To gather evidence related to this question, the evaluator attended the District Assessment Evaluation Team Training held at the Buros Center for Testing at the University of Nebraska on July 8-10, 2004. The purpose of this session was to train the individuals who subsequently scored the portfolios. The evaluator was unable to attend the Double Scoring Workshop held on September 16-17, 2004; thus, this evaluation cannot address any issues related to the quality of the scoring process beyond the scorer training workshop or to actual portfolio scores.

Dr. Chad Buckendahl, Co-Director of the Buros Institute for Assessment Consultation and Outreach, designed and, in partnership with Sue Graupner, facilitated the scorer training workshop. Dr. Buckendahl and other Buros staff are nationally-recognized experts in measurement and testing issues and have served as advisors to the NDE for several years on issues related to the development and scoring of the District Assessment Portfolios. The NDE considers the input of Buros staff when developing documents related to the portfolios.

In preparation for the scorer training workshop, Buros recruited 17 measurement professionals from across the country to serve as portfolio scorers. Most of these professionals were either former professors or graduates of the graduate program in Educational Psychology at the University of Nebraska and most are faculty members. At least half had served as portfolio scorers in previous year.

Two other groups attended the training workshop in addition to the prospective scorers. First, members of Nebraska's National Advisory Committee for Assessment² attended to observe the training process. Second, seven members of a cohort of Nebraska educators receiving advanced training in measurement

² Members of this committee are Jeri Benson (University of Georgia), Susan Brookhart (consultant), Brud Maxcy (Maine Department of Education), and Joe Willhoft (Tacoma Park Public Schools).

attended the training workshop. Three of these educators served as facilitators for the portfolio pre-submission workshops.

The first day of the scorer training workshop focused on a close review of the portfolio scoring rubric and specific requirements related to assessment quality criteria 1-4. Requirements related to assessment quality criteria 5 and 6 were reviewed on the morning of the second day. The same color-coded documents that had been used for the portfolio pre-submission workshops were used in this section of the scorer training workshop. Sue Graupner, who had facilitated the pre-submission workshops, also facilitated this section of the scorer training workshop. In total, scorers spent approximately nine hours reviewing criterion-related evidence requirements.

On the afternoon of the second day, scorers were trained on the use of the on-line portfolio rating form to be used for recording ratings and comments. Scorers then formed small groups of five to eight members and team-scored three portfolios. Actual 2004 portfolios were used in this section of the workshop and were selected from the pool of District Assessment Portfolios by Dr. Buckendahl and Ms. Graupner as exemplars of the performance levels in the scoring rubric.

All groups scored the same three portfolios in the same order. For each criterion, group members discussed the evidence in relation to the rubric until the group reached consensus on the rating. While groups were reviewing the training portfolios, Dr. Buckendahl and Ms. Graupner monitored the discussions and responded to scorers' questions on specific points of evidence. Once all teams had completed a portfolio, the large group was reconvened and the ratings for each criterion were discussed. In most cases, criterion ratings were identical across all groups. When groups differed in their ratings, these ratings were discussed by relating specific points of evidence from the portfolio to the rubric. These discussions resulted in consensus in all cases.

During the training portfolio reviews, several experienced scorers noted that the quality of the portfolios was higher than in previous years. In addition, a number of scorers commented during group discussions on the general superiority of the evidence for criteria 1-4 relative to the evidence for criteria 5 and 6. Scorers also indicated that it would be helpful to know whether the district whose portfolio was being reviewed received Title I funding, since that has implications for the level of information required for reporting.

Following the scorer training workshop, each scorer was assigned 15-17 portfolios to rate. Ratings were recorded using the on-line form and all were completed by August 13, 2004. Forty-five portfolios were rated as unacceptable and were automatically rescored by a member of the National Advisory Committee for Assessment. In addition, to evaluate consistency in ratings, 80 portfolios were selected for blind double-scoring. Scorers were reconvened for the double-scoring session on September 16-17, 2004. Results from this double-scoring process will be provided to the NDE by the Buros Institute and were not considered in the evaluation described here.

Evaluation Results. The evidence described above was considered in relation to the following question:

How did the criteria for scoring the District Assessment Portfolios align with the criteria that were communicated to local educators?

The underlying issue in this question is whether the scorers *were trained* to score portfolios using the same criteria that were used to guide local educators in the development of their portfolios. Therefore, both the content and the training process are relevant components of the evaluation results.

With regard to the content used for training, the criteria to which scorers were trained were highly aligned with the portfolio rating criteria that were communicated to local educators. This alignment was apparent in two respects. First, the same training documents were used and the same facilitator (Sue

Graupner) led the portfolio pre-submission workshops and the scorer training workshop. Second, throughout the three-day scorer session Dr. Buckendahl and Ms. Graupner grounded all discussions in the scoring rubric. Any questions that arose regarding individual ratings were answered using the rubric.

The training process also appeared to support the alignment between what had been communicated and how the portfolios were to be scored. The three-day training session was organized into two parts: (1) grounding in the criteria and scoring rubric and (2) practice scoring sessions. An extended amount of time was allowed for each part so that scorers were able to engage in the discourse and application necessary to solidify a common understanding of the rubric.

The second part of the training session, when scorers scored three actual portfolios, appeared to be critical to the shaping of scorers' conception of the rubric. As the scorers progressed through the scoring process for the three portfolios, the discussions grew shorter as groups reached consensus more quickly.

The automatic appeals and double-scoring session that occurred after the portfolios were rated also likely contributed to the meaningfulness of the portfolio ratings.

Two other features of the training workshop likely contributed to its overall quality and consistency over time. First, local educators who had facilitated or participated in the portfolio pre-submission workshops participated in the scorer training workshop. On a number of occasions, these educators were able to provide insights into how and why districts engage in assessment-related practices described in the portfolios. For example, these educators explained that most districts lack the capacity to conduct some higher level statistical analyses and that, in many cases, there were too few teachers to allow for independent item development and review within a given grade level and content area. Scorers discussed these constraints and came to the common understanding that this information should not be considered when scoring the portfolios. However, it could be helpful when writing comments designed to provide guidance for improving practice in the future.

Second, recruitment of experienced scorers and participation in the training workshop by members of the National Advisory Committee for Assessment likely contributed to consistency in the rating process over time. These experienced participants served as de facto peer facilitators in many cases, so that newer scorers' understanding of the rubric was developed in the context of the rubric as established in previous years.

Recommendations. The scorer training process supported the alignment between the scoring criteria and the criteria used to drive development of the portfolios by local educators. The following recommendations may contribute to the continued quality of these sessions:

- (1) Continue to involve at least some experienced scorers in the process to help ensure consistency over time.
- (2) Continue to involve local educators so that they can (i) help scorers to interpret information in the portfolios and (ii) learn about the scoring process so that they can subsequently inform local educators about this process.
- (3) Continue to involve both Dr. Buckendahl and Ms. Graupner, whose combined expertise and facilitation skills were critical to the quality of the training session.
- (4) Continue to structure the training session into two parts, the first involving intense consideration of the criteria and rubric and the second providing three opportunities to practice score actual portfolios.

- (5) As noted in relation to the first evaluation question, involve some experienced scorers in the portfolio pre-submission workshops to provide guidance and to support local educators' understanding of their audience when developing the portfolios.

Conclusions

The results of this evaluation indicate that the criteria that the NDE has developed regarding the quality of the District Assessment Portfolios are clearly communicated to local educators and are consistent with the criteria used to score the portfolios. Thus, for the 2003-04 school year, local educators have reason to trust that the portfolio criteria that have been shared with them are the same criteria that scorers will use to rate their portfolios.

With regard to future changes in the portfolio review process, the NDE must continue to build on the systems that have been established for communication and training. Over the past few years, the NDE has refined the documentation and training processes associated with the District Assessment Portfolios. These refinements have been informed by the results of external evaluations and by self-study, so the NDE has been proactive in its continuous improvement efforts in this area and has been responsive when areas for improvement have been identified. When major changes are considered necessary, as in the case of more rigorous expectations for some of the assessment quality criteria that will take effect in the 2006-07 school year, it appears that the NDE is careful to inform local educators well in advance of implementation.

The association between the NDE and the Buros Institute appears to be mutually beneficial and should be maintained. These partners may disagree on the correct course of action in some cases, which is not uncommon when academic and administrative agencies work together. However, the combined expertise in measurement and in the functioning and capacity of local education agencies can greatly benefit Nebraska educators by providing technically sound guidance that reflects practical limitations to implementation.